CGIAR Research Program on
Climate Change, Agriculture and Food Security (CCAFS)

Budgeting & Planning
Video Transcript

October 2013

These materials were produced for and with funding from the Climate Change Agriculture and Food Security Research Program of the Consultative Group on International Agricultural Research (CGIAR).

Budgeting & Planning for Data Management

This short presentation accompanies the document and the Terms of Reference for data management staff in this section of the support pack. The purpose of this video is to show you the typical data management tasks in a project the size of CCAFS and help you decide whether or not you need a data manager for your project.

Introduction

Data management is acknowledged as necessary but not always budgeted for explicitly within a project. Frequently data management is done by each researcher in a project independently from other researchers in the same project. This approach makes it difficult to apply the same principles and standards throughout the project. We encourage you to think of the resources you need for data management in terms of time and money.

Time & Budget

Consider the equipment you are likely to need – this might be equipment for data entry – you might consider hand-held devices for example. Think about backups – consider cloud storage or network drives – make sure these are budgeted for.

Ensure team members have the time to do the tasks well – don't forget to factor in time for dealing with data queries and time for training.

Data Management Role

The PI needs to decide whether or not to appoint a data manager – whatever the decision the data management tasks must be explicitly assigned – someone must have the responsibility to see that they are done.

Time and resources needed for data management depend on the specific project size and complexity.

Data Manager's Tasks

The data manager supports and contributes to the team's efforts to gather, clean, store and publish the research data and accompanying documentation. He or she has the responsibility to implement any agreements established by the research team with respect to data.

For example they might set up data and document storage facilities for the researchers to share their data. They might help in the creation of data entry systems including documenting the system and helping to train data entry staff. They might provide input to the design of data collection tools – considering the layout of a questionnaire from a data management point of view for instance can make it easier to design the data entry system which in turn would facilitate the computerisation of the data. The data manager would ensure Meta-data is produced, would help with data quality checks and would be responsible for archiving.

Example – CCAFS Data Management

As an example of the typical tasks that a data manager might be expected to do we have listed just some of the data management tasks carried out by the Statistical Services Centre for the CCAFS Baseline Surveys:

- Formatting the questionnaires for the household baseline survey for ease of completion and data entry
- Allocating variable names throughout the questionnaire
- Providing support in working with CSPro – creating demonstration videos
- Revising the data entry system in line with changes in the questionnaire
- Writing a data entry manual to be used in training and for reference
- Writing data checking guides – detailing how to produce frequency tables in CSPro, how to transfer the data to SPSS and how to do further checks in SPSS
- Helping to draft the analysis plan for the site reports
- Producing the SPSS syntax for the analysis plan
- Documenting the procedures for running the analysis and extracting the results
- Running quality checks on the data – this included things like making sure the land values were feasible and there was not more land used for growing crops or grazing than the farmer actually owned
- Merging the data sets from the 15 core sites – this included
- Consolidating the crop and livestock codes – there was a core set of codes but each site added codes specific to their region – these needed combining in the merged dataset
- Creating the Dataverse for archiving the baseline studies – this was a pilot run of using Dataverse and seems to have been successful
- Uploading the data and documentation to Dataverse – over time the structure of the Dataverse has changed to be more logical and further files have been uploaded as they have become available
- Setting up a data entry worksheet for the data from the mitigation questionnaires from the Village Baseline – this included entering and checking the data
- Checking site reports
- Re-running some of the analysis when corrections had been made to the data for example in one site the conversion factor for converting from the local land unit to hectares had been recorded incorrectly so farmers appeared to have much more land than was actually the case
- Creating training videos on how to complete parts of the questionnaire and how to use the data entry system – these included videos explaining how to record the land values – some sites had manually converted the data to hectares prior to data entry leading to problems when the analysis syntax did a further conversion
- Producing an assessment report on the quality of the household baseline data
- Translation of the data entry screens into Spanish for use in Nicaragua and Costa Rica – Also some of the screens were translated into Portuguese for use in Mozambique.
- Etc.

Are you really sure you don't need a data manager?

Summary

The debate about whether to appoint a data manager or leave researchers to do their own data management will no doubt continue. Some researchers have the capacity, time and inclination to do data management but in our experience they are the exception. In most cases support from a data manager who has explicit responsibility to deal with data issues is essential to achieving good quality data. In the support pack we have included terms of reference for both a data manager and a data technician or assistant.

Appendix I – CCAFS Data Management Support Pack

This document is part of the CCAFS Data Management Support Pack produced by the Statistical Services Centre, University of Reading, UK. The following materials are available in the pack:

0. Data Management Strategy
    a. CCAFS Data Management Strategy
1. Research Protocols
    a. Writing Research Protocols – a statistical perspective
    b. Preparation of Research Protocols – Good Practice Case Study
    c. What is a Research Protocol, and how to use one (Video & Transcript)
    d. Details of what a Research Protocol should contain (Video & Transcript)
2. Data Management Policies & Plans
    a. Creating a Data Management Plan
    b. Data Management Plan (Video & Transcript)
    c. Example Data Management Activity Plan
    d. Example Consent Form
3. Budgeting & Planning
    a. Budgeting & Planning for Data Management
    b. ToR Data Support Staff
    c. Budgeting & Planning (Video & Transcript)
4. Data Ownership
    a. Data Ownership and Authorship
    b. Template – Data Ownership Agreement
    c. CCAFS Data Ownership & Sharing Agreement
    d. Data Ownership & Authorship (Video & Transcript)
5. Data & Document Storage
    a. Creating and Using a DDS
    b. DDS Introduction – (Video & Transcript)
    c. DDS Organisation – (Video & Transcript)
    d. DDS Ownership – (Video & Transcript)
    e. Introduction to Dropbox – (Video & Transcript)
6. Archiving & Sharing
    a. Archiving & Sharing Data
    b. Data and Documents to Submit for Archiving – a checklist
    c. MetaData
    d. Archiving & Sharing (Video & Transcript)
    e. Metadata (Video & Transcript)
    f. CCAFS HBS Questionnaire
    g. CCAFS HHS Code Book
    h. CCAFS Training Manual for Field Supervisors

7. CCAFS Data Portals
    a. Portals for CCAFS Outputs
    b. AgTrials Summary
    c. CCAFS-Climate Summary
    d. DSpace Introduction
    e. Introduction to Dataverse (Video & Transcript)
    f. Creating a Dataverse (Video & Transcript)
    g. Dataverse Study Catalogue
    h. CCAFS Dataverse (Video & Transcript)
8. Data Quality & Organisation
    a. Data Quality Assurance
    b. Guidance for handling different types of Data
    c. Transition from Raw to Primary Data
    d. Data Quality Assurance (Video & Transcript)
    e. Guidance for handling different types of data (Video & Transcript)
    f. Transition from Raw to Primary Data (Video & Transcript)